PROCEEDINGS OF THE CALIFORNIA ACADEMY OF SCIENCES

Volume 56, Supplement I, No. 6, pp. 62–71, 2 figs. June 3, 2005

Species Numbers in Bacteria

Daniel Dykhuizen

Department of Ecology and Evolution, Stony Brook University, Life Sciences Building, Stony Brook, NY 11794-5245; Email: dandyk@life.bio.sunysb.edu

A modified biological species definition (BSD), i.e., that bacteria exchange genes within a species, but not usually between species, is shown to apply to bacteria. The formal definition of bacterial species, which is more conservative than the modified BSD, is framed in terms of DNA hybridization. From this I estimate there are a million species of bacteria in 30 grams of rich forest topsoil and propose that there will be at least a billion species worldwide.

Bacteria are a major component of the cellular life on Earth and are found everywhere from the top of mountains in Antarctica to the deep-sea vents. They are found in the deep subsoil, the open ocean and all over every surface of you. The refrain for undergraduates is that only about 10 percent of the cells moving with you are eukaryotic, the rest bacterial symbiotes. But because bacterial cells are so much smaller than eukaryotic cells, they make up only about 10 percent of your weight. (So, tomorrow, when you step on the scale you can subtract 10 percent off the scale weight because it is not really yours.)

Of the three great branches of cellular life, two are bacterial: the Eubacteria and the Archeae. The third branch is the Eukaryotes of which plants, animals, and fungi make up three kingdoms. The Eubacteria are divided into 40 kingdoms and the Archeae are divided into two kingdoms. The question I will try to answer in this paper is how many species of bacteria might there be. Before we can estimate the number of species of bacteria there may be in the world, we have to determine whether bacterial species are real entities and how they can be defined.

Bacterial Species

Bacteria are different from eukaryotes in an essential characteristic of life history. Bacteria are haploid and always reproduce asexually by fission. They have other mechanisms for sex, the transfer of genes between lineages, such as transformation, transduction, and conjugation.

The effects of sex in bacteria are different than in animals. In bacteria only a small fraction of the genome is transferred, unlike animals, where 50% of the offspring genome is transferred from the male during sex. Sex in bacteria can be between clones of the same species, closely related species, or distantly related species, whereas in animals sex is almost always limited to members of the same speeies. During sex in bacteria, pieces of genes or whole genes are transferred and replace the alleles present by homologous recombination, rather than form heterozygotes of the female and male alleles as in animals. Also, in bacteria, new genes can be added to the genome by non-homologous recombination.

The distinction between sexual and asexual as is made in plants and animals does not make sense for bacteria. The amount of sex, or lateral gene transfer between lineages, in bacteria can vary from very little to a lot. *Borrelia burgdorferi*, the causative agent of Lyme disease, is almost com-

pletely clonal, only transferring small pieces of DNA very rarely (Dykhuizen and Baranton 2001). In *Heliobacter pylori* (Suerbaum et al. 1998), the causative agent of stomach ulcers, and *Neisseria gonorrhoeae* (O'Rourke and Spratt 1994), the causative agent of gonorrhea, there is so much sex that alleles of different genes are in linkage equilibrium.

There is a certain flexibility in the bacterial genome as to what genes are present, even within a species. This part, which ranges from zero to about 20 percent of the genome (Ochman et al. 2000; Daubin et al. 2003a), is represented by genes that are transferred into the cell by non-homologous recombination and are fairly readily lost again. Belonging to this class of genes are DNA parasites like insertion sequences, transposons, lysogenic phage, small plasmids, and probably restriction modification systems. There are also genes that are involved in local adaptation to particular environments such as the genes in pathogenicity islands, antibiotic resistance genes in plasmids (mini chromosomes), genes for resistance to toxins and heavy metals, and genes and operons for various specialized functions. Generally, the genes in this pool of transients are either DNA parasites or genes that are locally adapted (Eberhard 1990). The genes that are generally useful to the organism across all the environments in which it lives are found in all strains and are usually associated with the major chromosome. Some of these genes are particular to a species or a small group of species, but most are widely found throughout bacteria. These genes, such as those involved in protein synthesis, DNA metabolism, energy generation and usage, etc., are often called housekeeping genes.

Some individuals (e.g., Gogarten et al. 2002) have been so enamored by this novel aspect of bacteria, the rapid and possibly widespread transfer of DNA across species, that they have suggested bacterial cells are simply holding vessels for the genetic variation available in the bacterial gene pool. The part of this pool seen in any particular cluster (species?) is simply the genes that are selectively useful given the current environment. Thus, bacterial phylogeny would represent ecology more than history.

I do not think this is true, but rather bacterial species can be defined in ways that are very similar to the way animal species are defined. I proposed in 1991 that bacterial species could be defined as a variant of the biological species definition (Dykhuizen and Green 1991). To restate and update that definition, I will define a species as a group of individuals where the observed lateral gene transfer within the group is much greater than the transfer between groups such that phylogenetic history is preserved when genomes are compared. Below I will illustrate what I mean by this definition. But before this, I want to make two comments. The first is that I think the similarity between species in bacteria and in animals occurs because species are real and caused by the same basic biology in both cases. Although it is still not clear what this basic biology is, just as it was not clear what basic biology the Linnaean hierarchical system of classification described before Darwin's theory of descent, I will present my idea of what this basic biology is later. The second involves the idea of "observed transfer" in the definition. The rates of observed transfer will depend upon both the rate of transfer and the selection for or against the strains containing these transfers. Generally, the rates of transfer will decrease with phylogenetic distance (Majewski and Cohan 1998) and the selection against the transferred DNA will increase with phylogenetic distance (Cohan et al. 1994), such that transfer is seen within species but not between species (see below). Of course, if the environment changes, as, for example, humans using tremendous amounts of antibiotics for the health and growth of themselves and their animals, even rare transfers of DNA from phylogenetically distance sources will be strongly selected if they provide resistance. The ability to incorporate DNA from outside the species is an advantage that generally distinguishes bacteria from animals. In animals and plants, incorporation of DNA from different taxa is unlikely, but can happen when associated with endosymbiosis.

The first example supporting a BSD involves *Escherichia coli* and *Salmonella*. It is well established that there is recombination between strains of *E. coli* (Dykhuizen and Green 1991; Guttman and Dykhuizen 1994). However, there is also evidence that there is little or no transfer between *E. coli* and *Salmonella* and none from more distantly related bacterial species into either of these two closely related species. Because this is not generally realized, I will describe the evidence in detail. Figure 1 shows the sequence distance between homologous genes of these two species plotted by a measure of codon bias (Sharp 1991). Codon bias refers to preferential use of certain codons over others even though all the codons code for the same amino acid. Highly used genes show more bias than less used genes. Highly used genes synthesize more protein, placing a larger demand upon the pools of charged tRNA. The preferred codons are the codons for tRNA types with the largest pools and the unpreferred codons are the ones with the smallest pools of tRNA. Thus, there is selection to use the preferred codons, which is stronger in highly used genes. Because the codon preference in *E. coli* and *Salmonella* is the same, genes with high codon bias will diverge less rapidly than ones with low codon bias (Fig. 1).

Like the dog that did not bark, Figure 1 is very clear in what it does not show. There are no points in the upper right and lower left sections (except the four open circles). If a gene had been transferred into either species from a more distantly related species, the distance between *E. coli* and the *Salmonella* copies would be much greater than expected for amount of codon bias seen. If there had been a recent transfer between *E. coli* and *Salmonella*, the gene copy that had been evolving in one species since the separation of the two species would be replaced by one from the other species and the distance would be too small given the codon bias. The circles in the figure are noted by the author as having too small a distance given the bias and are possible cases of lateral gene transfer between the species. However, he suggests that for the three open circles on the left, the mutation rate is lower because these genes are contiguous and near the origin, giving a smaller than expected distance, rather than these genes having transferred between the species well after the species split. This supposition has been strongly supported (Ochman 2003). The open circle on the right represents the *tufA* and *tufB* genes. These are very highly synthesized genes, expected to have strong codon bias. The position and phylogeny of the duplications make it very unlikely that the small distance was caused by recent lateral transfer rather than strong selective constraint (Sharp 1991). In conclusion, although there is strong evidence that there is recombination between lineages for housekeeping genes within *E. coli*, there is no evidence of gene transfer between *E. coli* and *Salmonella*. Thus, we can consider these as two distinct species by my definition of species. Using whole genome sequences, Daubin and collaborators (2003b) have shown recombination within *Escherichia coli* and *Chlamydophila pneumoniae*, but lack of recombination between at least seven species. Thus, it seems that my definition of species will be robust.

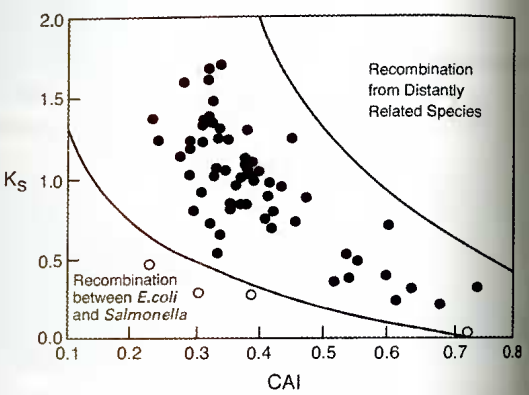

FIGURE 1. Plot of codon bias by sequence divergence between genes found in *E. coli* and *Salmonella*. CAI (Codon Adaptation Index) is a measure of codon bias. Ks is the number of substitutions per synonymous sites, where the distance is corrected for multiple hits using the Kimura two-parameter model (Kimura 1980). This figure is modified from Figure 1 of Sharp 1991. Codon bias explains 50% of the variation in synonymous site divergence. The open circles are the genes that might have been laterally transferred after the species split.

Salmonella enterica has been thought to have very limited recombinational exchange (Feil et

al. 2001, Maynard Smith 1995). This conclusion was based upon studying two examples of each of eight subspecies. To show recombination within a taxon, one needs at least three individuals. Thus, this test could not determine recombination within the subspecies. Brown et al. (2003) recently have shown that there is extensive recombination within one of the subspecies. Thus, the conclusion is that *Salmonella enterica* is likely to be eight separate species.

We suggested that if there is no genetic exchange between species, then all the gene trees should be congruent (Dykhuizen and Green 1991). Lawrence et al. (1991) showed gene tree congruence for three genes across a number of species in enterobacteria. More recently this work has been extended using whole genomes (Lerat et al. 2003). They found concordance for 203 out of 205 gene trees across 11 species. The two exceptions were inconsistent because of a single lateral gene transfer (LGT) event. The conclusion of Lerat et al. (2003) very strongly supports our definition of bacterial species: "Our analysis indicates that single-copy orthologous genes are resistant to horizontal transfer, even in bacterial groups subject to high rates of LGT" (p. 101).

The *Neisseria* are a group of species that are primarily commensals of the mucous membranes of mammals. There are a group of seven species that are commonly found associated with humans, six of which are found in the back of the mouth and the seventh, *N. gonorrhoeae,* in the urogenital tract. These bacteria are naturally competent for transformation throughout their entire life cycle and have high rate of LGT (Spratt et al. 1995). *N. meningitidis* and *N. gonorrhoeae* are closely related, with the DNA of coding genes >98% sequence identity. Although there is extensive genetic exchange within each species (Maynard Smith et al. 1993), there is little exchange between them (Vazquez et al. 1993), conforming to the Biological Species definition. This is presumably because they are physically isolated, living in different parts of the body. The six species living in the throat do exchange DNA even though they are more distantly related with sequence similarity among species ranging from 91% to 77%. Because of strong selection, pieces of the gene for the penicillin binding protein from two of the naturally penicillin resistant species have been incorporated into *N. meningitidis,* rendering it resistant to penicillin, even though the divergence is 14% for one species and 23% for the other (Spratt et al. 1995). Clearly, strong selection can incorporate genetic material from other species into the genome. However, what happens when there is no strong selection? When housekeeping genes are sequenced from these species, fragments from other species are present in many of the genes. These fragments are recent transfers that have not yet been purged by selection, because when the fragments are removed, the gene trees generally match (Maynard Smith 1995). If fragments had been incorporated in the past, the gene trees would not match. Thus, the fragments must have been selected against.

This is proven by an exception. The *adk* gene is scrambled (Feil et al. 1995). Why do we see extensive recombination in this gene and not in others? I think the answer is that *adk* has only one polymorphic amino acid across the genus whereas the other genes have many. Thus, I conclude that there is selection against fragments from other species because of amino-acid differences, i.e., the genome is co-adapted. This maintains species boundaries. Because of this general property of cellular life, the co-adapted gene complex, species definitions can be very similar for bacteria, animals and plants.

Number of Species in 30 Grams of Soil

Whereas there are many species definitions (e.g., Cohan 2002), I wish to use the formal definition of species in bacteria because it is both useful and conservative. It states, "The phylogenetic definition of species generally would include strains with approximately 70% or greater DNA-DNA relatedness and with 5°C or less Δt_m. Both units must be considered" (Wayne et al.

1987:463). Thus two strains are different species if less than 70% of the DNA will re-associate after having been melted to single strands. This criterion is required because up to about 30% of the DNA of a bacterial cell can be transient. Also, mismatches will decrease melting temperature of the re-associated DNA. For two strains to be different species, this decrease must be more than 5°C. This translates to 7–8% difference in DNA sequence (Caccone et al. 1988). Thus, by this conservative definition, *N. gonorrhoeae* and *N. meningitidis* would be considered the same species.

We can use this definition to estimate the number of bacterial species in a community. The rate of re-association of single stranded DNA with its homologue depends upon the concentration of the homologue. As the number of different fragments of DNA increases, the time becomes longer. Figure 2 shows the re-association of *E. coli* DNA and calf thymus DNA. The genome of a calf is larger than the genome of *E. coli*, consequently it takes much longer for the calf thymus DNA to re-associate than the *E. coli*. Re-association kinetics are measured in terms of the concentration of DNA in moles per liter (C_0) times the time in seconds (t). This is the Cot value. If the concentration of DNA is held constant, then the number of molecules of each unique sequence decreases as the genome size increases. For example, if the concentration of DNA is 12pg, a solution will contain 4000 copies of a genome of 0.003pg but only 4 copies of a genome of 3pg. In this example it will take on average 1000 times longer for the DNA in the large genome to find its homologue. The Cot value, when half the DNA is re-associated, gives an estimate of genome size. If we think of the bacteria from a natural community as a single species of bacteria, how large would its "genome" be? The number of species in the community can be estimated by dividing this "genome" by the average size of a bacterial genome.

Torsvik, Goksøyr and Dane (1990a) isolated 30 grams of top-soil from a beech forest north of Bergen, Norway. The soil contained 1.5×10^{10} bacteria per gram by microscopic observation. Less than 1% of these could be cultivated. After the eukaryotes and viruses were eliminated, DNA was extracted from the bacteria. The DNA was sheared, melted, and allowed to re-associate at a temperature that was 25°C less than the melt temperature. As seen in Figure 2, the re-association started at about the same time as the calf thymus DNA, which implies that the most common species is less than 1% of the population. The re-association at 50 percent is 10 times more complex than the calf thymus DNA. This gives an estimated complexity of 2.7×10^{10} bp. If you take the average genome size of soil bacteria at 6.8×10^6 bp (Torsvik et al. 1990b), which is a little larger than *E. coli*, you end up with an estimate of about 4000 common species. The rare species have still not re-hybridized. Actually the 4000 is an underestimate, because the re-association is 25 degrees below the melt temperature. If this re-associated DNA is melted, only 10 percent of the re-associated DNA fits within the definition of species (a Δt_m of less than 5°C). This suggests that we have underestimated the number of

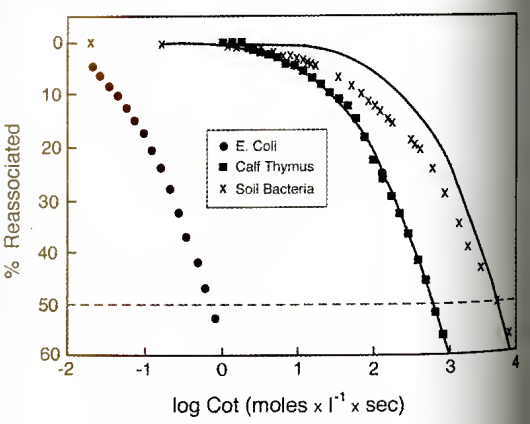

FIGURE 2. DNA rehybridization kinetics. This figure is a modification of Figure 3 of Torsvek, et al. (1990a). The *E. coli* genome is about 4.5 megabases and the unique DNA in cows is about 1,000 times larger as is human. Thus the initial hybridization of the soil bacterial DNA at same point that the calf thymus starts to re-hybridize suggests that the most common species in the soil is about one tenth of one percent of the total number. The line at the far right is the re-association kinetics for a homogenous sample like the calf thymus DNA, shifted so it has the same Cot as the soil bacteria at 50% re-association. This shows that the different species of DNA from the soil have very different frequencies.

species by ten fold. Using strains isolated from the soil, Torvick et al. (1990b) showed that the DNA hybridization gave a number of species ten-fold too low. So we end up with 40,000 common species.

There are always fewer common species than rare ones. If the species are ranked by number of individuals so that the most common species is first and the rarest species is last, then we can divide the species into two classes, each containing half the individuals in the sample. The ratio of the number of rare species to the number of common species gives us an estimate of the total number of species when we can only estimate the number of common species. Consider an example with 52 species. The two most common species make up half the number of individuals. Then the ratio of rare species to common species is 25 to 1. This is about average ratio found in the literature for animals and plants (Dykhuizen 1998). The ratio estimated from the data of Ruth Patrick (1968) on natural communities of diatoms is 25 if we use only the data from a single experimental box, but 35 if we combine the data from all eight boxes. Thus, as a first approximation, we will use a ratio of 25 to estimate the number of species in 30 grams of soil. This is 25 times 40,000 or a million species. There are caveats to this estimation. For example, a lot of the rarity might be in rare genes that are laterally transferring back and forth rather than in rare species. Thus, we need another way of estimating the number of species in this 30 grams of soil. This is provided by the work of Curtis and collaborators (2002).

Curtis et al. (2002) proposed using log normal species abundance curves to characterize bacterial communities. If the total number of bacteria and the number of the most common species are known and if it is assumed that the rarest species is present as a single individual, then the total number of species can be estimated. There are 5×10^{11} bacteria in 30 grams of soil and the most common species is between 1% and 0.1% of this number. Reading off Figure 4 of Curtis et al. (2002), the possible values for the numbers of the most common species give an estimate that span the value of a million species in this 30 grams of soil.

The Number of Bacterial Species in the World

We know very little about how many communities of bacteria there are and what the diversity may be within them. Although there may be a million bacterial species in 30 grams of rich soil in Norway, this might be one of the more diverse communities because it is in a nutritionally rich, structured and stable environment. I would expect similar soils (where I do not know what I mean by "similar") to have the same species, such that the number of species present in this type of soil worldwide would be about ten million in 3000 kilograms of soil (Curtis et al. 2002). The bacterial communities in sandy loam, sandy clay loam, loamy sand, and clay loam in England were tested by DNA-DNA hybridization and found to be different (Griffiths et al. 1996). There is some cross-hybridization between certain soil types. When there is some cross-hybridization, is this because the communities contain the same species in very different densities or is it because some of the species are shared between the communities, but most are different? Another way of asking this question is: "Are the rare species in one environment, common in another or are rare species rare and common species common?" The data of Patrick (1968) for a number of communities of protists in very similar experimental environments suggests the latter is true. Finley et al. (2002) suggests it is true for fresh-water protists. This might not be true of diatoms because the environment in the ocean is not structured as in soil and frewshwater lakes. However, I will assume it is also true for bacteria, i.e., rare species are rare and common, common, even in different soil types. Thus, the different soil types are likely to represent different communities of bacteria. How many soil types are there in the world that support different communities of bacteria? Do different plants on

the same soil type give different communities? We will need more data to answer these questions to be able to estimate how many species of soil bacteria there are in the world.

Curtis et al. (2002) suggested there are about two million species in the ocean. This is derived from estimating there are 163 species in a ml of seawater from the Sargasso Sea and then extrapolating this to two million in the ocean. However, the Sargasso Sea has very different bacterial community than Long Island Sound (Lee and Fuhrman 1990). Even within the Sargasso Sea, the communities at the surface and at 500 meters are different (Lee and Fuhrman 1990). Whereas there is no cross-hybridization between the Sargasso Sea and Long Island sound, there was some for the different communities at the surface and at 500 meters. Once again, is there cross-hybridization because the same species are present at the two depths, but in different ratios, or is there some species overlap, giving some cross-hybridization, but with most of the species different? If we assume the latter, then the estimate of Curtis et al. (2002) is probably a considerable underestimate.

In this symposium, Nancy Knowlton mentioned the large diversity of bacteria on corals and presented evidence for species specificity (Rohwer et al. 2002). We also know there is a community of bacteria in the deep sub-soils. Even at 500 m, the U.S. Department of Energy continues to find bacteria in their deep wells. Antarctic rocks contain bacteria that only metabolize three or four hours a year when the sun is directly on them; otherwise they are frozen. There are bacteria everywhere. Thus there must be many communities and consequently very many species.

From all this, my guess is there are a billion species and the more I get used to this number, the more I feel it is a gross underestimate. But for now, it is as far as my mind will go, given so little data. Thus, there are simply too many species of bacteria to count. Returning to Paul Ehrlich's argument, we need other measures of biodiversity, particularly for bacterial communities, than sampling and counting every species. Perhaps we can use DNA re-association measures or some method involving PCR to get estimates of the diversity. These measures can be used over time to look at community stability and ecosystem health. In an aside, it seems that PCR amplification directly from bacterial communities, sequencing these PCR products and estimating species number using rarification statistics from these sequences underestimates species diversity by about ten-fold.

Questions and Answers

How Many Named Species of Bacteria are There?

There are about 30,000 formally named species that are in pure culture and for which the physiology has been investigated. Species now are being defined by PCR amplifying ribosomal genes and sequencing. The criterion for defining species is that the ribosomal genes are at least 3% different. This method is probably is even more conservative than DNA/DNA re-association methods for defining species. We're probably defining species by ribosomal sequence at the level of genera or family.

What is Known About the Distribution of Bacterial Species Around the World?

Almost nothing. We know something about the biogeography of infectious diseases. Many of them are worldwide with very little population differentiation, like *E. coli*. Some of them are very specific to particular regions. So for example, the spirochete that causes Lyme disease is mostly found on the coastal plane of the east coast of the United States from the islands off Maine down to Maryland, and in the region of central Wisconsin, southern Minnesota and into the upper peninsula of Michigan. There is also a region in the central valley of California. There are low densities

of this spirochete in other areas, but it seems that the biotic cycle in these regions prevents high densities of both ticks and Lyme disease. Particularly along the northeast coast of the United States, it seems that the high density of Lyme disease matches the region of the last major glaciation. This correlation is probably because of the lower biodiversity (particularly of reptiles) after the glaciation. I must add one thing. The bacterial ecologists have a motto, which I don't think is completely correct, but does have some bearing. The motto is that is that everything gets everywhere, it's the environment that counts. So you find the same species in Yellowstone as you do in the hot springs in Iceland.

What will Happen as Biodiversity Decreases?

I'm speculating right now, but I think we're going to get more epidemics as biodiversity declines. What will happen is that the other species that are left will go up in numbers, and the organisms that infect them of course will go up in numbers. This means that there is a larger chance for them to jump to humans, which are a vast and untapped resource. We make the oil fields of Iraq look like a minor player as far as energy resources for bacteria go.

How Many Species of Bacteria are Found on Humans?

When I start my classes, I say that the bacteria on you are 90 percent of the cells that you walk around with. Only 10 percent of the cells that you walk around with are you. But they're much smaller. They are only 10 percent of the biomass. So you would only lose 10 percent of your weight if you got rid of all your bacteria. I don't think we really know how many species are found on humans. We're talking about 400-500 in your mouth, around the plaque. It seems that about 100 in your intestinal tract give you normal intestinal functioning. Germ-free mice don't function well. So that's not a very happy answer, but at least we're talking in the hundreds. Many of them are probably unique to humans.

Is Anybody Doing Similar Work Amongst Single Cell Eukaryotic Organisms?

Of this kind of measure within a system I don't think so. Single cell eukaryotic organisms are already much larger than bacteria. There's an attempt to do that picking out individual ones and looking at what species they would fall into. There's a lot of work trying to do the taxonomy of single cell eukaryotes. That's very exciting.

Acknowledgments

I thank the California Academy of Sciences for giving me the opportunity to present this material and to Nina Jablonski for patience. I thank Michael Feldgarden for proofreading and discussion. I was supported by NIH grant #GM60731. This paper is contribution #1136 from the graduate program in Ecology and Evolution at Stony Brook University.

Literature Cited

Brown, E.W., M.K. Mammel, J.E. LeClerc, and T.A. Cebula. 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. *Proceedings of the National Academy of Sciences USA* 100:15676–15681.

Caccone, A., R. DeSalle, and J.R. Powell. 1988. Calibration in the change in thermal stability of DNA duplexes and the degree of base pair mismatch. *Journal of Molecular Evolution* 27:212–216.

COHAN, F.M. 2002. What are bacterial species? *Annual Review of Microbiology* 56:457–487.

COHAN, F.M., E.C. KING, P. ZAWADZKI. 1994. Amelioration of the deleterious pleiotropic effects of an adaptive mutation in *Bacillus subtilis*. *Evolution* 48:81–95.

CURTIS, T.P., W.T. SLOAN, J.W. SCANNELLI. 2002. Estimating procaryotic diversity and its limits. *Proceedings of the National Academy of Sciences USA* 99:10494–10499.

DAUBIN, V., E. LERAT, AND G. PERRIERE. 2003a. The source of laterally transferred genes in bacterial genomes. *Genome Biology* 4:R57. <http://genomebiology.com/2003/4/9/R57>.

DAUBIN, V., N.A. MORAN, H. OCHMAN. 2003b. Phylogenetics and the cohesion of bacterial genomes. *Science* 301:829–832.

DYKHUIZEN, D.E. 1998. Santa Rosalia revisited: Why are there so many species of bacteria? *Antonie van Leeuwenhoek* 73:25–33.

DYKHUIZEN, D.E. AND G. BARANTON. 2001. The implications of a low rate of horizontal transfer in *Borrelia*. *Trends in Microbiology* 9:344–350.

DYKHUIZEN, D.E. AND L. GREEN. 1991. Recombination in *Escherichia coli* and the definition of biological species. *Journal of Bacteriology* 173:7257–7268.

EBERHARD, W.G. 1990. Evolution in bacterial plasmids and levels of selection. *Quarterly Review of Biology* 65:3–22.

FEIL, E., G. CARPENTER, AND B.G. SPRATT. 1995. Electrophoretic variation in adenylate kinase of *Neisseria meningitidis* is due to inter- and intraspecies recombination. *Proceedings of the National Academy of Sciences USA* 92:10535–10539.

FEIL, E.J., E.C. HOLMES, D.E. BESSEN, M.-S. CHAN, N.P.J. DAY, M.C. ENRIGHT, R. GOLDSTEIN, D.W. HOOD, A. KALIA, C.E. MOORE, J.J. ZHOU, AND B.G. SPRATT. 2001. Recombination within natural population of pathogenic bacteria: short-term empirical estimates and long term phylogenetic consequences. *Proceedings of the National Academy of Sciences USA* 98:182–187.

FINLEY, B J., E.B. MONAGHAN, AND S.C. MABERLY. 2002. Hypothesis: The rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist* 153:261–273.

GOGARTEN, J.P., W.F. DOOLITTLE, AND J.G. LAWRENCE. 2002. Prokaryotic evolution in light of gene transfer. *Molecular Biology and Evolution* 19:2226–2238.

GRIFFITHS, B.S., K. RITZ, AND L.A. GLOVER. 1996. Broad-scale approaches to the determination of soil microbial community structure. Application of the community DNA hybridization technique. *Microbial Ecology* 31:269–280.

GUTTMAN, D.S., AND D.E. DYKHUIZEN. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266:1380–1383.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111–120.

LAWRENCE, J.G., H. OCHMAN, AND D.L. HARTL. 1991. Molecular and evolutionary relationships among enteric bacteria. *Journal of General Microbiology* 137:1911–1921.

LEE, S., AND J.A. FUHRMAN. 1990. DNA hybridization to compare species composition of natural bacterioplankton assemblages. *Applied and Environmental Microbiology* 56:739–746.

LERAT, E., V. DAUBIN, AND N.A. MORAN. 2003. From gene trees to organismal phylogeny in prokaryotes: The case for the gamma-proteobacteria. *Public Library of Science Biology* 1:101–109.

MAJEWSKI, J., AND F.M. COHAN. 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus*. *Genetics* 148:13–18.

MAYNARD SMITH, J. 1995. Do bacteria have population genetics? Pages 1–12 *in* S. Baumberg, J.P.W. Young, E.M.H. Wellington, and J.R. Saunders, eds., *Population Genetics of Bacteria*. Cambridge University Press, Cambridge, England, UK.

MAYNARD SMITH, J., N.H. SMITH, M. O'ROURKE AND B.G. SPRATT. 1993. How clonal are bacteria? *Proceedings of the National Academy of Sciences USA* 90:4384–4388.

OCHMAN, H. 2003. Neutral mutations and neutral substitutions in bacterial genomes. *Molecular Biology and Evolution* 20:2091–2096.

OCHMAN, H., J.G. LAWRENCE, AND E.A. GROISMAN. 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.

O'Rouke, M., and B.G. Spratt. 1994. Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: Extensive genetic diversity within isolates of the same electrophoretic type. *Microbiology* 140:1285–1290.

Patrick, R. 1968. The structure of diatom communities in similar ecological conditions. *American Naturalist* 102:173–183.

Rohwer, F., V. Seguritan, F. Azam, and N. Knowlton. 2002. Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series* 243:1–10.

Sharp, P.M. 1991. Determinates of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: Codon usage, map position, and concerted evolution. *Journal of Molecular Evolution* 33:23–33.

Spratt, B.G., N.H. Smith, J. Zhou, M. O'Rourke, and E. Feil. 1995. The population genetics of pathogenic *Neisseria*. Pages 143–160 *in* S. Baumberg, J.P.W. Young, E.M.H. Wellington, and J.R. Saunders, eds., *Population Genetics of Bacteria*. Cambridge University Press, Cambridge, England, UK.

Suerbaum, S., J. Maynard Smith, K. Bapumia, G. Morelli, N.H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori*. *Proceedings of the National Academy of Sciences USA* 95:12619–12624.

Torsvik, V., J. Goksøyr, and F.L. Dane. 1990a. High diversity of DNA in soil bacteria. *Applied and Environmental Microbiology* 56:776–781.

Torsvik, V., K. Salte, R. Sørheim and J. Goksøyr. 1990b. Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. *Applied and Environmental Microbiology* 56:776–781.

Vazquez, J.A., L. de la Fuente, S. Berron, M. O'Rourke, N.H. Smith, J. Zhou, and B.G. Spratt. 1993. Ecological separation and genetic isolation of *Neissweria gonorrhoeae* and *Neisseria meningitidis*. *Current Biology* 3:567–572.

Wayne, L.G., D.J. Brenner, R.R. Colwell, P.A.D. Grimont, O. Kandler, M.I. Krichevsky, L.H. Moore, W.E.C. Moore, R.G.E. Murray, E. Stackebrandt, M.P. Starr, and H.G. Truper. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic Bacteriology* 37:463–464.